Errors to test

R2 SCORE and RMSLE

RNN:

Epochs 20 :

Training RMSLE = 1.10

valifysadauypojn RMSLE = 1.10

Kaggle = 2.33 R2 = 0.60

Epochs 50 :

Training RMSLE = 1.07

val;idatijhn RMSLE = 1.10

R2 = 0.61

Epochs 10:

Training RMSLE = 1.11

validation RMSLE = 1.11

R2 = 0.59

Decision Tree:

Max depth 14 & random state 1 & 75% percent of training data used

Kaggle = 2.291 public

**Analysis**

In summary, most methods produce predictions with high levels of accuracy, with the worst being LSTM with an RMSLE of 5.616, and the best being Lightbgm with an RMSLE of 1.067. For some of the models, such as the decision tree model, there were certain performance and memory issues (Due to time complexity of O (m n^2), where m is the size of the data and n is the number of layers) that prevented them from producing nearly as accurately as they could and extra hyperparameter tuning could have been performed. Whilst the RNN model is not as accurate as Lightbgm, the short running time and low memory usage makes it an effective model to use if speed is a priority. As the two best models were the neural networks, it can be concluded they were the most effective method for this task.

In terms of errors, the greatest potential for errors within this project was the dataset itself. Some columns within the training dataset had to be removed entirely (Columns with greater than 50% missing data). Columns with less than 50% missing data had the missing data points replaced with the median of that column. Whilst it would be ideal if all the data was available, by keeping as many columns as possible, this led to optimal feature selection to enhance the accuracy of the neural network models.

Each model has its own weaknesses that could also cause errors. LSTM and Lightbgm are prone to overfitting and RNN frequently have gradient vanishing and exploding problems. Decision trees are unstable meaning that if there is a small change in the training data, the optimal tree may have large changes.

Sources:

<https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>

Jiang Su and Harry Zhang, A Fast Decision Tree Learning Algorithm, University of New Brunswick, NB, Canada

|  |  |
| --- | --- |
| **Model** | **RMSLE (From Kaggle Submission)** |
| Decision tree | 2.391 |
| Lightbgm | 1.067 |
| RNN | 2.069 |
| LSTM | 5.616 |
| Baseline Average | 3.069 |